高校图书馆读者数据的监管与增值利用
——以2017年北京大学阅读报告为例

吴亚平 别立谦 周春霞 赵飞 汪聪
北京大学图书馆 北京 100871

摘要：［目的/意义］提出读者数据利用框架，以有助于引导图书馆更规范地对读者数据进行系统的收集、整理、保存和挖掘，实现其增值利用。［方法/过程］引入数据监管和基于数据生命周期的数据服务拓展理念，建立基于生命周期的读者数据监管框架，并基于此框架，以《2017年北京大学阅读报告》为例，分析如何面向图书馆和读者进行读者数据监管。［结果/结论］读者数据监管一方面可以帮助图书馆明确读者来访及借书规律、核心读者群体、资源利用情况，合理优化各类资源和服务；另一方面通过向读者推荐阅读资源和服务，帮助读者提高自身阅读质量，使其充分地利用图书馆，从而为读者数据监管和增值利用的规范化、深入化发展提供有益参考。

关键词：高校图书馆 读者数据 数据监管 数据增值
分类号：G250
DOI：10.13266/j.issn.0252－3116.2019.18.008

1 引言

图书馆业务的开展和读者天然紧密地联系在一起，读者的需求和喜好是图书馆优化的方向和动力。大数据专家徐子沛在《数文明》新书发布会上提到，数据正在引领人类高清晰社会，重塑文明[1]。可以理解为数据可以让我们更了解社会。同样，在图书馆界，数据也将帮助我们更清晰地看到读者的需求和图书馆的资源、服务效用。大数据的环境下，图书馆对读者数据价值的重视提升到了新高度。国际图联趋势报告中明确提出，图书馆员需要对版权、数据保护和隐私的含义有更好、更广泛的理解[2]。图书馆已逐步开设了数据馆员岗位，目前其岗位名称有数据馆员、数据服务馆员、数据可视化分析馆员、数据管理馆员、数据存储馆员与研究数据专家等[3]。基于读者特征数据和行为数据的创新应用已经渗透到多种服务中，如基于对读者借书、翻阅报刊、浏览电子资源等行为数据的分析，为资源建设、资源布局、空间优化等提供决策支持。近些年，在智慧图书馆、个性化图书馆的发展理念下，图书馆更注重读者数据的分析和展示，除了日常的挖掘与利用外，更是推出阅读报告、大数据报告、个人阅读账单、实时数据展示屏等多种形式的数据阅读服务，这些都是读者数据监管和增值利用的实践。笔者在相应工作的基础上提出读者数据监管框架，以期推动读者数据被更规范、更深入地增值利用。

2 相关研究

2.1 数据监管是实现数据增值的重要途径，图书馆应担当好“监管人”角色

“curation”一词最早应用于西方文艺遗产保护方面，博物馆学中称之为“策展”，即策划、筛选并展示，“curator”指对脆弱而又贵重的博物保存和展览活动进行构思、组织、管理的人员[4]。在E-Science研究环境下，“Digital Curation”于2001年被首次提出，英国数据策展中心（Digital Curation Center，DCC）将其定义为对数据进行整个生命周期的管理、维护、保存，以实现数据增值的所有活动[5]。2002年，微软首席研究员、计算机图灵奖获得者J. Gray提出“data curation”，并指出短暂数据具有不可替代、不可重建的特性，是必须要保存的，比如每一天的气温，若不记录保存，则无法复

作者简介：吴亚平（ORCID：0000－0002－4242－2434），馆员，E-mail：wuyp@lib.pku.edu.cn；别立谦（ORCID：0000－0003－4725－908X），研究馆员；周春霞（ORCID：0000－0003－1440－9969），副研究馆员；赵飞（ORCID：0000－0001－7395－7485），馆员；汪聪（ORCID：0000－0003－0750－8563），馆员。
收稿日期：2018－12－12 修回日期：2019－03－05 本文起止页码：63－69 本文责任编辑：易飞

原[6]。国内学者通常将其翻译为数据策管、数据监护、数据监管、数字保护和数据掌管等,来表达其在对数据进行选择、组织、存储的基础上实现数据复用、增值的含义,从而维持数据的可生存能力、可呈现能力和可理解能力[7]。此处的数据包括数字资产、数字数据、数字科研数据、原始资料、数据集等[8]。本文借鉴最常见的 3 种译法之一,将其翻译为"数据监管"。读者数据作为短暂数据的一种,对图书馆资源和服务的优化有重要作用,必须及时地保存、管理并加以增值利用。

数据增值是对专门学科领域的数据进行广泛的收集,之后对数据进行验证和关联,通过可视化技术展示关联结果,从而在保护数据的同时实现数据增值[9]。数据监管的过程包含了数据增值的关键步骤,是数据增值的重要途径。图书馆作为信息资源的拥有者和提供者,除了收集、整理和挖掘信息,更要将所得信息产品按照用户习惯或一定的主题思路,艺术性地展现在用户面前[10],担当好"监管人"的角色。在图书馆逐步开展的大数据阅读服务中,也要求数据具有较强的科学性、实时可用性、生命周期活动可控性和数据管理动态可维护性,数据监管正是读者阅读服务质量(quality of service,QOS)的重要保障[11]。

2.2 基于读者数据的阅读服务和数据决策更日常化,群体也更加细分

年度阅读报告、个人阅读账单、阅读清单、实时数据展示屏等都是读者数据监管的体现。年度阅读数据报告逐渐成熟,截至 2018 年 1 月 26 日,已有近百所图书馆发布了 2017 年阅读数据报告[12]。从视角的不同,可以分为综合性阅读报告和个性化阅读报告,综合性阅读报告以图书馆为中心,整体性、群体性强,可对资源利用和读者阅读情况形成全局的认识;个性化的阅读报告以读者为中心,描述读者个性化的阅读行为,有更强的读者存在感和参与感。除年终统计外,读者数据报告也体现出常态化、精细化的特点,如沈阳师范大学图书馆[13]发布暑假阅读数据、黑龙江大学图书馆[14]发布寒假阅读数据,沈阳师范大学图书馆还推送月度的热门外借书单[15],众多图书馆甚至推出实时数据展示大屏。分析的群体从整个读者群体细化到了新生、毕业生、校友等,如北京大学图书馆面向新生推出"新生特辑|当萌新遇见图书馆"[16],面向毕业生推出"书・时光"毕业纪念卡套装,支持在线生成和分享电子版阅读纪念卡[17];厦门大学图书馆面向毕业生推出的"圕・时光:给毕业生的一份礼物"[18];四川大学图书馆面向校友推出的"毕业生的不二情书'一封家书'"[19]等,这些个性化的阅读数据展示,记录了读者的阅读足迹,极具收藏价值,深受读者欢迎。随着数字媒体显示技术的进步及液晶屏、LED 屏等显示设备的发展,实时数据显示屏成为了图书馆宣传服务的新方式。如上海图书馆利用可视化技术和多媒体展现技术,直观、实时地展示业务统计数据,既可以为图书馆决策提供支持,又是一种大数据应用的阅读推广方式[20]。还有学者提出,未来甚至可以专门设置读者个性化区域,基于读者的借阅历史建立个性化知识模型,读者刷卡后即可显示可能感兴趣的图书及其基本信息、在架情况等[21]。大数据阅读服务作为一种新的服务形式,是读者数据监管的表现形式,即基于一个既定目标,搜集处理数据,挖掘数据内涵,并借助数据可视化技术艺术性地展示出来,从而指导图书馆更好地优化资源和服务,指导读者更好地提升自身阅读质量。

2.3 数据可视化技术日渐成熟,在数据监管中举足轻重

数据可视化是数据监管的关键步骤,如今人们已经对单调保守的讲述方式失去兴趣,期待更加直观、高效的信息呈现方式[22]。人类 80% 以上的信息通过视觉系统获得[23],"眼见为实"也正说明人们更信赖视觉获取的信息。数据可视化提供了易于阅读与优美的视觉体验,更高效地传达了数据内涵,如今已应用到了多个领域。包括位置、形状、方向、颜色、纹理、灰度等级、尺寸等在内的可视化变量也在不断扩展[24]。作为一种高效的信息展示形式,可视化技术的应用渗透到了图书馆的多个业务场景:①在馆藏资源引导方面,以直观的图形或图像的方式展示图书馆的资源布局情况,多用于新生引导和日常引导;②在资源揭示方面,提取资源的属性特征,作为揭示内容的维度,如世界数字图书馆以时间轴和地图的可视化形式向读者展示世界各地文化的原始资料,促进了国际文化的交流[25];③在资源检索方面,实现了馆藏资源关联表示可视化、馆藏资源查询过程可视化、馆藏资源检索结果可视化等[26];④在业务统计方面,可视化的展示有助于更高效的管理决策,如厦门图书馆业务数据可视化平台集中展示读者信息和馆藏使用情况,为图书馆文献采购、数字化馆藏建设、馆藏剔旧等决策提供了科学依据[27]。此外还有基于 Timeline 的图书馆大事记可视化[28]、引文分析[29]等方面。如今,图书馆更加看重数据可视化,并设置相关的岗位提供更专业的服务。如杜克大学、加州大学伯克利分校、哈佛大学均设置了数据可视化分析馆员,要求具备数据提取、清理、转换、分

析技术，掌握数据可视化技术等能力[30]。

综合各种数据可视化技术，可将其分为面向低维度数据的柱状图、折线图、饼图等标准2D/3D技术，以及适用于多维度、大数据集的几何投影技术、基于图像技术、面向像素的技术、分层技术等新技术[31]。数据可视化工具也发展迅速，包括：excel等入门级工具；E-Charts、D3.js等开源工具；R、Gephi等专业工具；Tableau等商业工具；等等。随着不断升级发展，可视化工具具有更强的易操作性、实时性、动态性和丰富的展现形式、支持多类型数据格式等特点。可视化技术可以更直观、快捷、准确地捕捉数据内涵，在读者数据的监管中将起到关键作用。

3 基于生命周期的读者数据监管框架及实践

数据监管包含对数据进行整个生命周期的管理、维护、保存，以实现数据增值的所有活动。数据生命周期是从数据产生、加工、发布到再利用的一个循环过程，基于数据生命周期的数据服务拓展可划分为数据初加工、数据再加工、知识抽取三类[32]。笔者结合数据监管内涵及基于数据生命周期的数据服务拓展，提出包含三个阶段、八大模块的图书馆读者数据监管框架。

3.1 数据初加工阶段

数据初加工阶段是数据增值的基础和保障，其内容和框架如图1所示：

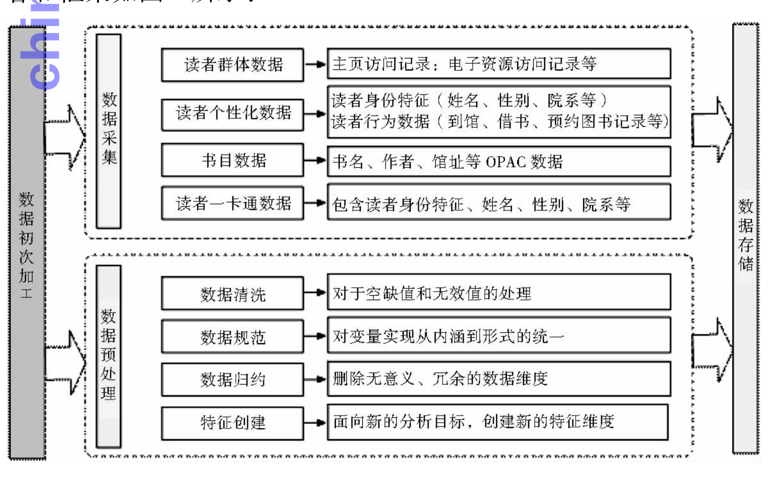


图1 数据初加工阶段的监管框架

数据采集阶段目的是全面地搜集读者数据。读者数据包括4个方面：①电子资源访问等读者群体数据，有助于从整体角度，客观地体现发展趋势、发现热门资源；②借书记录等含有读者ID的个性化数据，是开展个性化数据服务产品的基础；③OPAC系统中的书目数据，更详细地描述图书特征，是对读者行为数据中图书属性的重要扩展；④读者一卡通数据，是对读者行为数据中读者身份信息的重要扩展。

完成数据采集后的数据预处理模块是数据增值利用的重要前提，主要包括4个部分：①数据清洗，表现在通过第三方数据源对缺失数据进行填补，如通过学校一卡通数据对借书记录里读者院系是“部门未知”的记录进行匹配、补缺，最大程度消除信息不确定性，以及对无关数据进行删除，如删除借书记录中夹杂的装订记录；②数据规范，对同一属性值的不同表现形式进行统一，如将同一院系的全称和简称统一为全称；③数据规约，即对冗余的属性进行删除，如读者院系编号和院系名称，只保留院系名称字段即可；④特征创建，根据分析目标创建新的特征，如基于院系字段创建学部字段，从而探讨不同学部的读者借书特征。数据存储则贯穿始终，既要存储原始数据，也要存储处理过程和结果数据。

3.2 数据再加工阶段

数据再加工阶段是数据增值的重要体现，内容和框架见图2。

基于数据初加工阶段得到的规范的异构、多源数据，首先要建立数据连接，增加数据关联。主要包括两个方面：一方面是基于读者ID整合读者一卡通数据、借书记录、进馆记录、预约记录、续借记录等读者行为数据，从而拓展读者特征，全面地还原读者在图书馆的行为；另一方面是基于图书ID整合读者行为数据与OPAC书目数据，拓展读者行为数据中的图书特征。在建立多维数据连接后，可从时间、资源类型等分析维度进行一般性统计分析，得出发展趋势，发掘热门作者及资源等，也可进行多属性间的交叉分析，分析得出男女最喜爱的作者，不同学部读者利用资源的差异等。数据分析与可视化展示相辅相成，数据分析赋予可视化展示价值与意义，可视化展示帮助从数据中提取价值和知识[33]。在利用数据可视化的手段时，首先要明确可视化的原则和希望展现的目标，并基于此选择合适的可视化方法和工具，使最终的效果具有“油墨比”高效、视角清晰、组合维度、对比恰当、动态可交互等特征，从而充分利用人类的视觉带宽，提升用户对数据的思考能力和理解效率[34]。

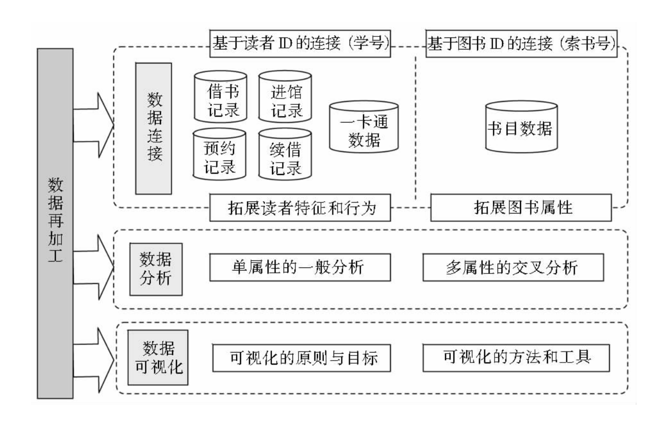


图 2　数据再加工阶段的监管框架

3.3　知识抽取阶段

知识抽取阶段是数据增值的最高等级，内容和框架如图 3 所示：

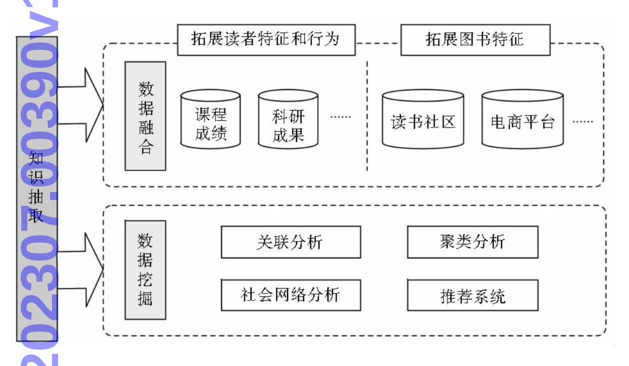


图 3　知识抽取阶段的监管框架

知识抽取是知识服务的前提，大数据环境下的知识服务须同步爆发式数据增长与社会化趋势，将视角瞄准碎片化信息、用户行为、用户关系，并聚焦由此产生的实时数据、非结构化数据及机器数据，调用内部、外部以及公共信息，从而做出前瞻性的数据判断[35]。读者数据监管更要利用好图书馆内部、校内其他部门和网络公开数据。除自身数据外，一方面要整合校内课程成绩、科研成果等读者数据，挖掘更多数据关联，例如美国大学与研究图书馆协会发布报告表明，使用图书馆服务（借阅、数据库访问、馆际互借等）多的学生往往取得了更高水平的学业成功（如课程成绩、GPA 等）[36]；另一方面要多渠道获取书目数据来拓展图书特征，书目数据相当于一本书的名片，是用户了解书籍的窗口，单一的书目数据源不足以多方位描述一本书的特征，Web2.0 的资源共享理念带来了越来越多的网络资源，Amazon 等图书电商网站和豆瓣等书评社区中所包含的评分、标签、评论等读者行为数据是对图书馆 OPAC 书目数据的重要补充。

通过数据融合得到读者大数据集是知识抽取的基础，数据挖掘技术是知识抽取的重要手段。目前数据挖掘技术主要应用于读者分析研究、资源建设优化、智能化服务、信息自动化处理、个性化服务等多方面[37]。读者数据监管服务需要更深层次地挖掘数据内涵，通过关联挖掘，得到阅读兴趣相似的读者以及内容相似的资源；通过聚类分析，针对更具体的读者和资源群体，开展精细化服务；通过社会网络分析，找到关键的读者和资源，以最优宣传策略达到阅读推广的目标；通过推荐系统，实现精准的个性化阅读推荐等。

4　读者数据监管结果与展示

《2016 年北京大学阅读报告》[38]是北京大学图书馆首次推出的年度阅读报告，对阅读资源和服务进行了系统的梳理，对读者利用图书馆的特征进行了分析、展示。《2017 年北京大学阅读报告》[39]在其基础上进行了内容和形式的优化，提升了规范性、全面性、美观性，是读者数据监管与增值利用的重要成果。在第 3 节监管框架和流程的基础上，《2017 年北京大学阅读报告》主要从面向图书馆优化服务和面向读者优化阅读两方面进行监管。

4.1　面向图书馆优化服务的监管

面向图书馆优化服务的监管，以明确读者来访及借书规律、核心读者群体、资源利用情况为目标，从而帮助图书馆合理优化各类资源和服务。

4.1.1　*从读者来访看*　读者来访不再限于实体到馆，主页访问、微信关注等来访渠道都起到了重要作用。12:00－12:59 是读者平均到馆人次最高的时段，不同时期的日均到馆人次有差异，考试周实地到馆人次明显增加，而假期到馆人次明显减少（见图 4），图书馆可基于读者密度，更科学地调配人力等资源，调整服务。微信公众号中的馆藏检索栏目点击次数和点击人数均为最多，可见微信公众号不仅是宣传的利器，其功能性同样受欢迎，是资源发现的重要入口之一。

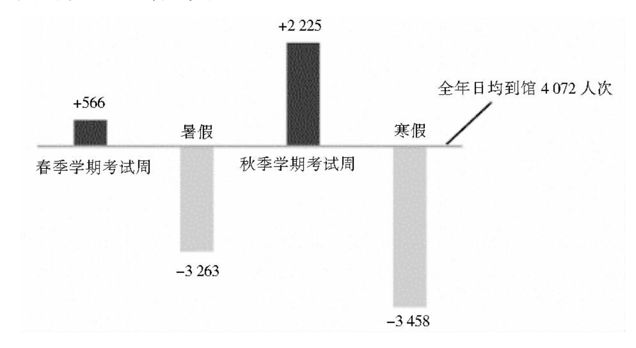


图 4　2017 年读者不同时期日均到馆人次差异

4.1.2　从资源利用情况看　数据库、多媒体等电子资源以及资源发现系统、自助服务等更便捷的资源搜索、获取方式越来越受欢迎。2013－2017年,“未名学术搜索”、电子资源全文下载篇次、多媒体资源在线检索与点播频次均持续上升,电子资源和相关服务越来越重要。阅读设备更新换代较快,Kindle系列的阅读器目前最迎合读者需求,在资源预约榜中遥遥领先。同时,自助服务也担当了重要角色,如自助借书册次占50.34%,超过了人工借书。分馆在文献保障体系中起到了重要作用,面向全校读者开放的服务的分馆数量不断增加。2017年分馆图书外借量总计达8万余册次,信息科学技术学院分馆的外院系读者外借量占比排名第一,达50%。针对不同文献采取的不同馆藏方式与读者借书实际情况相符,体现出合理性,可以看出开架借书的比例高于闭架借书、储存馆借书、馆际互借方式(见图5),2017年读者通过闭架借书和储存馆借书比例明显高于2016年,这是由于2017年图书馆为配合东馆改造工程进行了资源的重新布局,开架空间受到压缩,数据也反映出了闭架借书、储存馆借书、馆际互借方式借书量的增加。

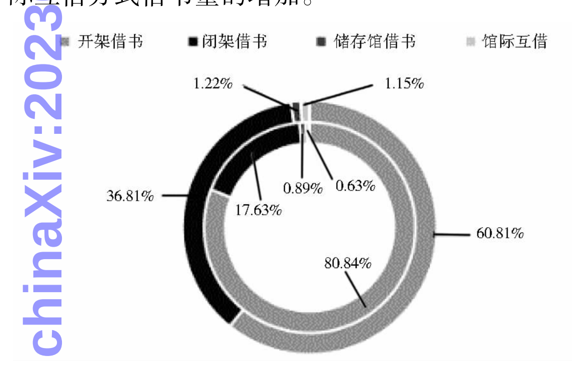


图5　2016(内环)/2017(外环)年读者不同方式借书量占比

4.1.3　从读者特征与阅读习惯看　人文学部、社会科学学部的读者是使用图书馆空间和实体资源较多的读者群体;人文学科类图书的跨学科渗透率更强,在各学部读者的借书所属学科占比中都占据了较大的比重。在到馆方面,人文学部和社会科学学部的读者分别占到馆读者的30%和21%,二者共占据了半壁江山;在借书方面,二者之和达到了66%(其中人文学部44%,社会科学学部22%)。外国语学院的到馆总人次和借书总册次最多,马克思主义学院的人均借书量和有借书行为的读者占比最高。读者的跨学科阅读趋势十分明显(见图6),可以看出每个学部的读者在阅读本学科的图书外,对其他多个学科都有涉猎。

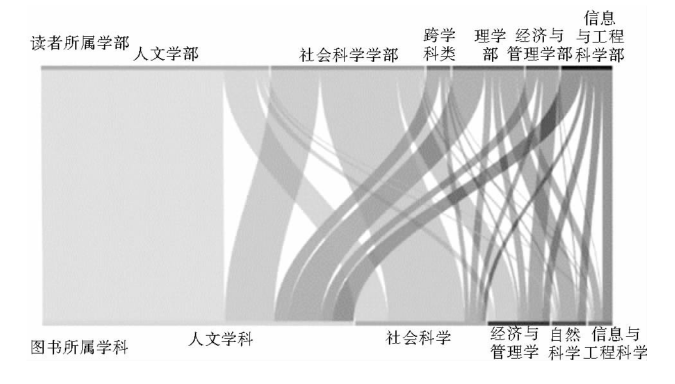


图6　2017年读者所属学部与图书学科类别匹配度情况

4.2　面向读者优化阅读的监管

面向读者优化阅读的监管,以向读者推荐阅读资源和服务为目标,从而帮助读者提高自身阅读质量,更加充分地利用图书馆。首先向读者整体介绍图书馆的资源、服务与开放时间;其次展示热门外借的中外文图书、热门预约图书、热门电子书、热门教参书等热门资源,男/女读者最喜爱的作者(见图7)等,向读者推荐热门资源;最后以时间轴的形式展示2017年的阅读活动,包括人文讲座、电影讲座、北京大学首届阅读马拉松活动、“书声迢递”世界读书日系列活动、文化工作坊活动,吸引读者回看讲座视频,并关注来年的系列活动,激发了读者的学习兴趣,也对阅读资源进行了推广。



图7　2017年最受男/女读者喜爱的前50位作者排名

对读者数据的监管与增值利用的实践主要集中在数据初加工和数据再加工阶段。通过面向图书馆优化服务和面向读者优化阅读两方面的监管,一方面有助于图书馆对资源和服务进行全景扫描,为图书馆资源建设、服务布局、人员调配等方面提供决策支持;另一方面有助于读者全面认识图书馆的资源和服务,并发

掘优质、热门资源,从而提升自身阅读质量。不足之处是知识抽取阶段的实践仍然较少,实时的数据驱动、精细的服务划分、个性化的推送将是阅读报告未来优化和应用的方向。

5 结语

现代信息技术从本世纪初开始推动着图书馆向智慧图书馆方向演进与发展,人工智能时代的到来,也必将驱使当前智慧图书馆的服务满足社会服务需求,具有智能特征,进而向智能图书馆方向发展[40]。读者数据监管与增值利用,是大数据环境下增强图书馆智能特征的重要手段。笔者基于阅读报告的实践,初步建立了数据生命周期的读者数据监管框架,并结合实践梳理了数据初加工、数据再加工两个阶段的内容,面向图书馆和读者分别进行了监管,形成了决策参考。然而目前在知识抽取阶段缺乏深入、规模性的应用实践,对数据挖掘技术和可视化技术的应用也处于初级阶段,读者数据监管的广度和深度亟待提升。未来数据监管的深入实践与应用,要整合馆内、校内、可网络公开获取的数据,广泛拓展读者特征和图书特征,通过数据挖掘和大数据分析技术,得出数据规律和知识,搭建实时、动态的数据展示平台,提升数据的可解读性、可理解性和可观赏性,最终实现数据增值的目标。

参考文献:

[1] 涂子沛:数据正在引领高清晰社会,重塑文明[EB/OL].[2018-08-31]. https://blog.csdn.net/r6Auo52bK/article/details/84001471.
[2] 国际图联趋势报告——2016 新进展[EB/OL].[2018-08-31]. https://trends.ifla.org/files/trends/assets/trend-report-2016-update-zh.pdf.
[3] 蒋丽丽,陈幼华,陈琛. 国外高校图书馆数据馆员服务模式研究[J]. 图书情报工作, 2015, 59(17):56-61.
[4] 丁培. 数据策展与图书馆[J]. 图书馆学研究,2013(6):94-98.
[5] BEAGRIE N, POTHEN P. Digital curation: digital archives, libraries and e-science seminar[EB/OL].[2018-08-31]. http://www.ariadne.ac.uk/issue30/digital-curation.
[6] GRAY J, SZALAY A S, THAKAR A R, et al. Online scientific data curation, publication, and archiving[J]. Proceedings of SPIE-the International Society for Optical Engineering, 2002, 4846:103-107.
[7] 李文文,陈雅. 国内外 Data Curation 研究综述[J]. 情报资料工作,2013(5):35-38.
[8] 王海宁,丁家友,聂云霞. Digital/Data Curation 的概念与翻译研究[J]. 图书馆杂志, 2018(1):8-18.
[9] 朱易佳,刘盈盈. 数据增值:以欧盟数据网站为实例的新探索[J]. 图书馆学研究, 2016(12):50-55.
[10] 蔡屏. 网络环境下未来图书馆的服务定位——信息策展人[J]. 情报资料工作,2011(5):61-64.
[11] 马晓亭. 图书馆大数据监护系统的构建——以生命周期理论为视角[J]. 图书馆建设, 2014(12):31-33.
[12] 近百所图书馆 2017 年数据发布[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/IWvyBu4kCUUL1Cf5SBWyOQ.
[13] 图书馆暑假小数据[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/eDZV9RNtkCr-x7JwjPJQNA.
[14] 图书馆 2018 年寒假阅读数据[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/SQzGSdujbxF_-Pl-NEIqvQ.
[15] 6 月沈阳师范大学图书馆图书借阅排行榜[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/Vqys-rXkQSXxY-qIO-QyEtg.
[16] 新生特辑|当萌新遇见图书馆[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/nM6zguTX8DLOQGJqtuIV_A.
[17] 图书馆 2018 毕业季系列活动开始啦! [EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/uEtER7GMudrEvqm2Zu815Q.
[18] 圕・时光:给毕业生的一份礼物[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/Otz8ZWYbBxOMELjVFrGKzQ.
[19] 毕业数据大放送|你有一封来自图书馆的不二情书[EB/OL].[2018-08-31]. https://mp.weixin.qq.com/s/5syQ7n-oX-vySJ9dp8lpGFg.
[20] 杨佳,赵斌,宋歌笙,等. 上海市公共图书馆 2014 年阅读报告暨创新型数据服务汇编[J]. 图书馆杂志,2015,34(4):113-130.
[21] 吴蛟鹏. 大数据时代图书馆读者行为分析及应用[J]. 图书馆研究,2017,47(2):108-111.
[22] 赵斌. 数据可视化在上海图书馆数据展示服务中的应用[J]. 图书馆杂志,2015,34(2):23-29.
[23] 杨峰. 从科学计算可视化到信息可视化[J]. 情报杂志,2007,26(1):18-20.
[24] 杨达. 数字图书馆信息可视化的研究框架[J]. 沈阳大学学报(社会科学版),2005, 7(3):127-130.
[25] 世界数字图书馆[EB/OL].[2018-08-31]. https://www.wdl.org/zh/.
[26] 邱均平,余厚强,吕红,等. 国外馆藏资源可视化研究综述[J]. 情报资料工作, 2014(1):12-19.
[27] 张肖回. 基于数据仓库技术的图书馆业务数据可视化监控平台——以厦门图书馆为例[J]. 现代情报, 2013, 33(4):150-153.
[28] 胡晓. 基于 Timeline 的图书馆大事记可视化[J]. 图书情报工作,2013,57(s1):286-287,147.
[29] 李运景,侯汉清. 引文分析可视化研究[J]. 情报学报,2007(2):301-308.
[30] 蒋丽丽,陈幼华, 陈琛. 国外高校图书馆数据馆员服务模式研究[J]. 图书情报工作, 2015,59(17):56-61.
[31] 坎塔尔季奇. 数据挖掘:概念、模型、方法和算法[M]. 王晓海,

吴志刚,译. 北京:清华大学出版社, 2013.

[32] 师荣华,刘细文. 基于数据生命周期的图书馆科学数据服务研究[J]. 图书情报工作, 2011,55(1):39 -42.

[33] 杨利军,高军. 图书馆个性化服务中的大数据可视化分析与应用研究[J]. 现代情报, 2015, 35(7):68 -72.

[34] 黄玺磊. 大数据的最后一公里——数据可视化技术[J]. 中国金融电脑,2017(2):37 -43.

[35] 蒋勋,刘喜文. 大数据环境下面向知识服务的数据清洗研究[J]. 图书与情报,2013(5):16 -21.

[36] Documented library contributions to student learning and success: building evidence with team-based assessment in action campus projects [EB/OL]. [2018 -08 -31]. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/issues/value/contributions _ y2.pdf.

[37] 俞锦梅. 数据挖掘在国内图书馆应用领域研究综述[J]. 图书与情报,2015(2):137 -141.

[38] 2016 年北京大学阅读报告[EB/OL]. [2018 -08 -31]. http://www.lib.pku.edu.cn/portal/cn/zhanlan/yuedu2016-1.

[39] 2017 年北京大学阅读报告[EB/OL]. [2018 -08 -31]. http://www.lib.pku.edu.cn/portal/cn/zhanlan/yuedu2017.

[40] 陆婷婷. 从智慧图书馆到智能图书馆:人工智能时代图书馆发展的转向[J]. 图书与情报,2017(3):98 -101,140.

作者贡献说明:

吴亚平:设计论文框架,撰写和修订论文;

别立谦: 提出研究思路,修订论文;

周春霞: 调整论文框架,修订论文;

赵飞: 处理数据,修订论文;

汪聪:采集数据,修订论文。

Reader Data Curation and Value-added in University Libraries: A Case Study of the 2017 Peking University Reading Report

Wu Yaping Bie Liqian Zhou Chunxia Zhao Fei Wang Cong

Peking University Library, Beijing 100871

Abstract: [**Purpose/significance**] The paper put forward the reader data utilization framework to guide the library to systematically collect, organize, save and mine reader data more systematically and realize its value-added utilization. [**Method/process**] The paper introduced data curation and data lifecycle-based data service development concepts, and built a lifecycle-based reader data curation framework. Based on this framework, the paper took the practice of the 2017 Peking University Reading Report as an example, curating for libraries and readers. [**Result/conclusion**] Through the curation of reader data, on the one hand, it helps the library to clarify the readers' access and borrowing rules, core reader groups, resource utilization, and rationally optimize various resources and services. On the other hand, by recommending reading resources and services to readers, readers can improve their reading quality and make full use of libraries. It provides a useful reference for the standardized and in-depth development of reader data curation and value-added utilization.

Keywords: university library reader data data curation data value-added